SNOMED Clinical Terms®
UK Clinical Edition

SNOMED CT Query Table

A SNOMED CT data derivative
Reduces query execution false negatives occurring due to concept inactivation
**Document Management**

**Revision History**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Summary of Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>18-Nov-2013</td>
<td>First draft</td>
</tr>
<tr>
<td>0.9</td>
<td>21-Mar-2014</td>
<td>Review draft</td>
</tr>
<tr>
<td>1.0</td>
<td>01-Apr-2014</td>
<td>First release</td>
</tr>
<tr>
<td>1.01</td>
<td>01-Oct-2014</td>
<td>Second release</td>
</tr>
<tr>
<td>1.02</td>
<td>01-Apr-2015</td>
<td>Third release</td>
</tr>
<tr>
<td>1.04</td>
<td>01-Oct-2015</td>
<td>Fourth release</td>
</tr>
<tr>
<td>1.05</td>
<td>01-Oct-2016</td>
<td>Sixth release – reformat to NHS Digital document template</td>
</tr>
<tr>
<td>1.06</td>
<td>01-Apr-2017</td>
<td>Seventh release – MySQL demonstrator added</td>
</tr>
<tr>
<td>1.07</td>
<td>01-Oct-2017</td>
<td>Eighth release – minor clarifications of text</td>
</tr>
<tr>
<td>1.08</td>
<td>01-06-2019</td>
<td>Eleventh release – product changes due to derivation from RF2</td>
</tr>
<tr>
<td>1.09</td>
<td>01-10-2019</td>
<td>Twelfth release</td>
</tr>
<tr>
<td>1.10</td>
<td>12-08-2020</td>
<td>30.0.0 Release – clarifications in response to feedback</td>
</tr>
</tbody>
</table>

**Future Revisions Due**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
</table>

**Reviewers**

This document must be reviewed by the following people:

<table>
<thead>
<tr>
<th>Reviewer name</th>
<th>Title / Responsibility</th>
<th>Date</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denise Downs</td>
<td></td>
<td>21-Mar-2014</td>
<td></td>
</tr>
</tbody>
</table>

**Approved by**

This document must be approved by the following people:

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Title</th>
<th>Date</th>
<th>Version</th>
</tr>
</thead>
</table>

**Glossary of Terms**

<table>
<thead>
<tr>
<th>Term / Abbreviation</th>
<th>What it stands for</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSCIC</td>
<td>Health and Social Care Information Centre (NHS Digital)</td>
</tr>
<tr>
<td>IHTSDO</td>
<td>International Health Terminology Standards Development Organisation. Organisation responsible for the management of SNOMED CT.</td>
</tr>
</tbody>
</table>
Related Documents

<table>
<thead>
<tr>
<th>Document Title</th>
<th>Document URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK Concept history substitutions table</td>
<td>doc_UKSnomedCTHistorySubstitutionTable_Current-en-GB_GB1000000_20181001</td>
</tr>
</tbody>
</table>

Document Control:
The controlled copy of this document is maintained in the HSCIC corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

1 SNOMED® and SNOMED CT® are registered trademarks of the IHTSDO® (www.ihtsdo.org)
1 Introduction

This document describes the SNOMED CT Query Table, a computed derivative of the UK Clinical Edition of SNOMED CT. When used as an extension to the traditional Transitive Closure table (and not a substitute for it), the product aims to facilitate querying over SNOMED CT data where either or both the data being queried over, or the query specifications themselves, contain SNOMED CT concept identifiers with ‘inactive’ status. The use cases for the product therefore overlap with those of the SNOMED CT History Substitution Table product, which is also one of the inputs to the computation used to produce the Query Table. The technical report ‘SNOMED CT: Inactive Content’ provides further detail on the issue and use cases.

1.1 Document Scope

This document provides an outline of:

- the rationale for creating the new product
- the intended use cases
- the method of construction and known content issues

The document does not discuss clinical governance, safety or acceptance testing issues.

1.2 Product Scope

The Query Table product addresses ‘false negative’ reporting issues that may otherwise occur but only in relation to concepts found in the SNOMED Clinical Terms® International and UK Clinical Editions.

It does not currently cover similar issues relating to concepts from SNOMED Clinical Terms® UK Drug Edition.

1.3 Audience

Suppliers and clinical users of clinical information systems already based upon, or that process SNOMED CT or that are planning future systems with these capabilities.

Secondary users of data derived from SNOMED CT systems, including large data repositories, health networks, research units, research networks, CSUs and Commissioning Groups, Public Health Observatories

NHS programmes and organisations contracting or commissioning related activity involving either suppliers or users listed above.

1.4 Product Status and History

From October 2018, This document plus the table it describes constitute a product at the ‘Supported Product’ stage as defined within the Product Lifecycle².

‘Supported Product’ status means:

1. Both the release format specification of the product and the method of its content preparation shall remain fixed indefinitely unless a significant safety risk is identified that cannot be mitigated without changing them. Where changes are deemed necessary to improve a product then a formal consultation procedure will be undertaken which may include some or all parts of the product development process

² http://systems.hscic.gov.uk/data/uktc/snomed/governance/lifecycle.pdf
and may include an option for parallel running (i.e. support for both existing and new specification)

2. NHS Digital commits to continue to support, maintain and publish the product against that fixed specification indefinitely, subject to the considerations above or proper product termination

3. Quality assurance may be ongoing but the product is approved for deployment in live clinical systems, subject to standard safety assessment procedures associated with deployment of any product into a live environment

4. The commitment to release against a stable specification does not preclude continued parallel evolution of the specification and consequent development of improved variants which may or may not be considered as new products

Approval to prepare and release the SNOMED CT Query Table as a draft product was obtained from the Terminology Edition Committee on December 12th 2013. The first release occurred on April 1st 2014, as a Technology Preview and the committee approved the product as a Supported Product on 1st October 2018.

The algorithm used to generate the table was improved at the third (April 2015) release to include measures to similarly mitigate the risks of false reporting in respect of descendants of 363743006|Navigational concept| and Situations with the default context. The resulting table was as a consequence significantly larger than for the earlier releases.

A further improvement was made at the April 2017 release, and a sample implementation/demonstrator also added.

The additional mitigations added from the April 2015 release (in respect of navigational and default context content) are removed from the October 2017 release, pending detailed clinical safety review in respect of what are necessarily currently only partial and thus somewhat unpredictable mitigations for the risk of ‘false negative’ reporting.

From June 2019 the algorithms for generating the History Substitution and Query tables have been extensively modified to take RF2 rather than RF1 data as their only input. This change results in significantly different outputs for both tables, particularly in respect of the computed date on which concepts became inactive and therefore which side of the Query Table’s 2005 cut off for inclusion they fall (see section 4).

1.5 Enquiries

Content error reports, or requests for clarification or additional functionality, should be made to information.standards@hscic.gov.uk with ‘SNOMED CT Query Table’ in the subject line.
2 Background

2.1 Concept inactivation in SNOMED CT

SNOMED CT, like Clinical Terms Version 3, is a dynamic terminology: not only can new content be added, but existing content may be edited or made 'inactive' where necessary. SNOMED CT partitions all its concepts into those that are 'Active' and those that are 'Inactive'. In RF2, this initial binary partitioning is achieved through the active field in the sct2_Concept table; RF1 releases employ a ConceptStatus metadata value on each concept within the sct1_Concept table.

'Active' currently further subdivides into 'current' or 'pending move', whilst 'Inactive' has seven subdivisions: 'retired (no reason)', 'duplicate', 'erroneous', 'ambiguous', 'outdated', 'limited', and 'moved'. In RF2, this further partitioning is represented through the membership of the 900000000000489007|Concept inactivation indicator attribute value reference set (foundation metadata concept)|. In RF1, different ConceptStatus values appear in the sct1_Concept table; whether the concept is fundamentally ‘Active’ or ‘Inactive’ is never explicitly recorded in RF1 but rather only implied through this ConceptStatus value.

Refer to the SNOMED CT Technical Implementation Guide for detail on concept status.

In this document concepts are referred to as either active or inactive.

Although ‘retired’ is often synonymous in common usage with ‘inactive’, this document treats ‘retired’ as a reserved word for use only to refer to the concept status value ‘1’ in the SNOMED CT standard (RF1).

In addition to changing their ConceptStatus, inactive concepts also lose all modelled attributes on inactivation. In particular, this includes losing any IS_A relationships that previously positioned the concept within SNOMED CT's primary taxonomic hierarchy; inactive concepts exist instead in a separate, very flat, taxonomic tree.

For example, until October 2011, 308130008|Recurrent chest infection| was an active concept in the main taxonomy, classified below 275498002|Respiratory tract infection| as shown in Figure 1. Note that 195746005|Recurrent chest infection| also existed, but was already inactive (being a duplicate of the active 308130008) and, accordingly, was not classified as a hierarchy descendent of 275498002|Respiratory tract infection|, although SNOMED CT content does state that the two are related in some other, non-hierarchical way (blue horizontal lines).

From October 2011, \textit{308130008|Recurrent chest infection} also became inactive: ‘chest’ often means the entire thorax whilst, in use, ‘chest infection’ \textit{usually} means more specifically the lungs. Both \textit{308130008} and \textit{195746005} are therefore now (re-)classed as ‘ambiguous’, and \textit{308130008} also moved outside the main taxonomy: neither is a subtype of \textit{275498002|Respiratory tract infection}. Again, however, SNOMED CT content \textit{does} state that some other non-hierarchical type of relationships exist (blue horizontal lines).

2.2 Effect of concept inactivation on reporting

The movement of inactive concepts out of the main taxonomy means that, if SNOMED CT has been deployed into systems and regularly updated with new release data, and data has been recorded using codes like \textit{308130008} that have since become inactive, then data processing will behave predictably but inappropriately in retrieval or reporting contexts, unless systems implement an appropriate mitigation.

Where query specifications have been written to exploit only SNOMED’s main taxonomy (the vertical black arrows in Figures 1 and 2) to dynamically compute the codes that should be in the result set each time a query is run then, in most cases, such inactive data will not be returned by otherwise appropriate query specifications. This will occur even though the same coded content was previously correctly returned when the same query was run against an earlier version of SNOMED CT. Clinically, concept inactivation will therefore manifest as new ‘false negatives’: patients who are no longer returned by a particular query but where the clinical expectation is that they should be.

The SNOMED CT Query Table is intended to be an appropriate and sufficient mitigation.

2.3 SNOMED CT historical relationships

The Query Table is constructed as a computed derivative of other information about inactive concepts that is represented within the SNOMED CT release data. In addition to demonstrating how inactive concepts move out of the main taxonomy, Figure 1 and Figure 2 (above) also show how most inactive concepts are linked to alternative active concepts, within the main taxonomy, by means of non-taxonomic ‘historical relationships’. In the figures, these links are represented by the horizontal blue lines. In Figure 1, for example, the inactive \textit{195746005} has a SAME\_AS relationship to \textit{308130008}, indicating that the two concepts are semantically equivalent. But only one of the pair sits within the traditional retrieval taxonomy.

2.3.1 UK Substitutions Table

By postprocessing all such historical relationships it is possible to construct a ‘substitutions table’ listing all currently inactive concept identifiers and, for each, all possible substitutions by an active concept \textit{(provided such substitutions exist – see section 5)}. NHS Digital therefore provides a pre-computed Substitutions Table for the UK Edition in the same Subpack as the Query Table itself.
2.3.2 SNOMED Query Table

The Substitutions Table could be used as part of an update process to be performed at each new release of SNOMED CT: all historically captured data, all queries and all other design artefacts involving references to SNOMED CT concepts might be examined for occurrences of newly inactivated concepts and any found are ‘annotated’\(^4\) with their designated active substitute concept listed in the Substitutions Table, to be used for the purposes of future retrieval. However, this search-and-replace operation can be computationally prohibitive to perform in live systems, assuming that it is even known where all design artefacts that could involve newly inactive SNOMED codes are located.

The SNOMED CT Query table offers an alternative solution: whenever a taxonomic query is executed, inactive codes are logically (but not also physically) updated with their active substitute. This logical transform is achieved by extending the standard Transitive Closure table (itself normally a required component of environments that support taxonomic querying) to also include additional rows computed as the cross-product JOIN between the Substitutions and Transitive Closure tables.

The primary intended effect of the Query Table, therefore, is that a search that should return the active concept will also return all inactive equivalents, and vice versa.

There are, however, two important caveats or limitations of using the Query Table:

- inactive concepts for which no active substitute could be identified remain as a residue of potentially ‘unreportable’ false negatives; the true clinical risk of this effect is however believed to be extremely small (see s5).

- Using the Query Table may find new false positives over and above those that would be returned if the Query Table were not used at all. This effect can be partly controlled by proper use of the ISAMBIGUOUS column (see s0)

\(^4\) Medicolegally, the originally entered code selection must always persist and never be changed. For taxonomy-driven reporting purposes, however, it will often be more correct to use a different code to be either computed logically on the fly (eg. by using the Query Table at run time) or stored physically as further slots in the information model (eg. by lookup on the Substitutions table).
2.4 Related SNOMED CT Content Issues

In addition to the behaviour of inactive concepts, described above, SNOMED CT contains two other areas of content that can also give rise to false negative reporting. Between April 2015 and April 2017 these risks were also but only partially mitigated by the Query Table; this functionality was withdrawn from the October 2017 release:

2.4.1 Navigational concepts

At the time of writing, 727 active SNOMED CT concepts reside in a parallel hierarchy below 363743006|Navigational concept (navigational concept)|. Concepts previously contained within the primary taxonomy are moved here if the set of concepts normally given as their descendants can only be determined by reference to:

- the lexical properties of their descriptions:
  - 273855008|Assessment scales Sw-Va (navigational concept)|
  - 302302007|Enzymes A - L (navigational concept)|
  - 302303002|Enzymes M - Z (navigational concept)|

- logical properties that cannot be represented in SNOMED CT’s description logic:
  - 106220000|Genetic, molecular AND/OR cellular finding (navigational concept)|
  - 362976009|Finding / observation / clinical history (navigational concept)|

- opinion rather than objective fact
  e.g. 350084001|Alternative medicines (navigational concept)|

As with inactive concepts, the concepts in this navigational subhierarchy have no descendants and also lack all clinically expected ancestors.

Unlike inactive concepts, there is no modelled relationship between these concepts and some substitute in the main taxonomy, because no such substitute exists.

In most cases, the concepts found in the subhierarchy represent the projections into SNOMED-space of abstract grouper concepts found within the antecedent terminologies, principally SNOMED RT and Clinical Terms Version 3 (CTV3).

For some of those Navigational Concepts derived from CTV3, it is possible to determine the direct parents and children from the main SNOMED taxonomy by reference to NHS Digital’s Data Migration Products.

Between April 2015 and April 2017 the SNOMED Query Table therefore also contained rows to reposition some Navigational Concepts back into the main taxonomy. These measures were, however, limited: some Navigational Concepts remained outside the Query Table scope (mainly those deriving originally from SNOMED RT) whilst those that were treated were not necessarily guaranteed to always be ‘classified’ entirely correctly. Although overall the number of false negatives relating to Navigational Concepts was significantly reduced, and the number of new false positives very low, these additional rows were removed from the October 2017 release pending a more detailed clinical safety assurance. No mitigation may be safer than a partial mitigation, especially if partial mitigation can be misconstrued as complete mitigation.

2.4.2 Situation with explicit context

The SNOMED CT Context Terminological Model provides a within-terminology mechanism that allows findings to be characterised by ‘context’, including: whether findings are present or absent and procedures performed or not performed; whether the coded statement relates

---

5 See s6.2.3.5 in http://www.ihtsdo.org/fileadmin/user_upload/doc/en_gb/tig.html
to the subject of the record or to a relative and whether in the past, present or future. This mechanism is mostly intended to support contextualisation of postcoordinated expressions but it is also embodied in the small subhierarchy of precoordinated concepts below 243796009|Situation with explicit context|, wherein currently resides some 6,700 concepts (findings, family history and procedures) with a variety of contexts.

The context model further states that:

*When a SNOMED CT code appears in a record without any explicitly stated context, that code is considered to have a default context. The default is "soft" in that it can be over-ridden by information carried in the structure of the record or its information model.*

*The default context for a clinical finding code implies that the finding has actually occurred (vs. being absent), that it applies to the subject of the record (the patient), and that it is occurring currently or occurred at a past time that is given by a date - time record linked to the code.*

*The default context for a procedure code implies that the procedure was completed, that it was performed on the subject of the record (the patient), and that it was done at the present time or in the past at a time that is given by a date - time record linked to the code.*

In other words, for every code without context found in the findings or procedures hierarchy, a semantically equivalent expression implicitly also exists in which the default context has been made explicit. All such equivalent expressions will be found in the Situations subhierarchy; some– a very small minority of what is possible - have already been instantiated there:

- 267093002|Has a red eye (situation)| is equal to 246676003|Injection of surface of eye (finding)|
- 162057007|Nausea present (situation)| is equal to 422587007|Nausea (finding)|
- 287752006|Bursa sutured (situation)| is equal to 88626007|Suture of bursa (procedure)|
- 432729003|Cortisone injection given (situation)| is equal to 432861006|Injection of cortisone (procedure)|

Note that the reverse is not true; not every pre-coordinated, contextualised concept from the situations subhierarchy has a semantically equivalent non-contextualised one in the findings or procedures hierarchy. Many concepts below 243796009|Situation with explicit context| assign a different context to the default implied by non-contextualised content:

- 274823004|Clonus absent (situation)| is not equal to 36649002|Clonus (finding)|
- 408566000|Echocardiogram declined (situation)| is not equal to 40701008|Echocardiography (procedure)|
- 160350005|Family history: Squint (situation)| is not equal to 22066006|Strabismus (disorder)|

The true ‘contextually equivalent’ pairs are important to identify because, although both exist within the full SNOMED CT taxonomy, a taxonomic query for one does not normally also return the other. Similarly, many credible queries can be formulated that would return one and not both.

From April 2015 to April 2017 the SNOMED Query Table therefore also contained the additional ancestor/descendent rows that were valid for truly ‘context equivalent’ pairs within the findings, procedures and situations hierarchies, where it was possible to identify a pair.

Thus, when using the Query Table:

- 162057007|Nausea present (situation)| subsumes 422587007|Nausea (finding)| and vice versa
- 287752006|Bursa sutured (situation)| subsumes 88626007|Suture of bursa (procedure)| and vice versa
- but:
- 274823004|Clonus absent (situation)| still does not subsume 36649002|Clonus (finding)|
- 160350005|Family history: Squint (situation)| still does not subsume 22066006|Strabismus (disorder)|
How contextually equivalent pairs are identified

To explicitly state the default context for a concept in the Findings subhierarchy, the following modelling pattern is used:

243796009|Situation with explicit context|:
{246090004|Associated finding|=<finding>,
408729009|Finding context|=410515003|Known present|,
408732007|Subject relationship context|=410604004|Subject of record|,
408731000|Temporal context|=410512000|Current or specified|}

…and, as a result, all and only all such concepts should be autoclassified under 373573001|Clinical finding present (situation)|.

Similarly, the following modelling pattern explicitly states the default context for a procedure:

243796009|Situation with explicit context|:
{363589002|Associated procedure|=<procedure>,
408730004|Procedure context|=385658003|Done|,
408732007|Subject relationship context|=410604004|Subject of record|,
408731000|Temporal context|=410512000|Current or specified|}

…and should result in all and only all such concepts being autoclassified under 443938003|Procedure carried out on subject (situation)|

However, some procedure and finding content from the UK extensions may not be classified under these two groupers, because UK Extension content is not currently built using an autoclassifier. As a consequence, the algorithm to identify pairs of Findings or Procedures with an equivalent SNOMED CT Situation content with the default context is necessarily more complicated: all concepts with Fully Defined status under the more general 243796009|Situation with explicit context (situation)| concept are inspected. A deep structural analysis is performed on each to determine whether its modelling is an example of one or other design pattern shown above. If it is, the equivalent Finding or Procedure is determined to be the modelled value of the 363589002|Associated procedure| or 246090004|Associated finding| attribute, as appropriate.

Note, however, than many situation-with-default-context concepts exist as Primitive (ie not Fully Defined). For these, the non-contextualised equivalent from the Findings or Procedures hierarchies can not reliably be automatically identified, though they may exist. The Query Table therefore does not address these undetectable equivalent pairs.

Thus, for example, although the primitive situation-with-default-context concept 720008007|Prostate cancer care review done| is explicitly (and correctly) modelled with a relationship pointing at its equivalent Procedure:

720008007|Prostate cancer care review done|:
{363589002|Associated procedure|=720007002|Prostate cancer care review (procedure)|,
408730004|Procedure context|=385658003|Done|,
408732007|Subject relationship context|=410604004|Subject of record|,
408731000|Temporal context|=410512000|Current or specified|}

…other situation-with-default-context primitives have no such associated_procedure relationship modelled at all:

819311000000107|Weight management plan completed|:
…whilst still others have a modelled associated_procedure relationship but not to the direct equivalent concept, but rather to some more abstract ancestor of it:

522381000000101|Ear care information leaflet given|:
{363589002|Associated procedure|=71388002|Procedure|,
408730004|Procedure context|=385658003|Done|,
408732007|Subject relationship context|=410604004|Subject of record|,
408731000|Temporal context|=410512000|Current or specified|}

Despite existence of 699118000|Provision of ear care information leaflet (procedure)|

**October 2017 update**

Because of the limitations in the underlying modelled content outlined above, the Query Table can currently achieve only a modest and somewhat unpredictable reduction in the number of false negatives relating to ‘context equivalent’ pairs. Pending a more detailed clinical safety assurance, therefore, these rows are absent from the October 2017 and future releases.
3 Summary of Use Cases

The following use cases are illustrated by reference to the real historical changes to SNOMED CT content that were summarised in Figure 1 and Figure 2:

3.1 EPR Instance Data contains inactive conceptId
After October 2011, a taxonomic search for all patients with 50417007|Lower respiratory tract infection| or any of its subtypes should find patients coded with either 308130008|Recurrent chest infection| or 195746005|Recurrent chest infection, as well as those coded with 448739000|Recurrent lower respiratory tract infection|.

3.2 Query contains inactive conceptId
After October 2011, a taxonomic search for all patients with 308130008|Recurrent chest infection| should find patients coded with 448739000|Recurrent lower respiratory tract infection|.

3.3 Inactive conceptIds in Query and EPR instance data
After October 2011, a taxonomic search for all patients with 308130008|Recurrent chest infection| should find patients coded with either 195746005|Recurrent chest infection| or 448739000|Recurrent lower respiratory tract infection| or 308130008|Recurrent chest infection|. 
4 Design and Content Optimisations

A full SNOMED CT Query Table, containing all additional rows relating to all concepts ever inactivated or otherwise with an equivalent, would be a very large table indeed; in excess of 25 million rows. The UK SNOMED CT Query Table therefore makes certain assumptions that originally allowed for a much more modest table size of just under 6 million rows though both improvements to the algorithm and the ever increasing number of inactive concepts it needs to handle have caused the size to grow; the operational goal is to permit selective optimisations of the table that yield improvements in runtime compute performance but without not significantly reducing utility in most care settings.

Additional rows that would be valid in the full table are currently omitted if they derive from an inactive concept and any of the following are true:

- The inactive concept was inactivated prior to 1st January 2005, and so there should be very little or no SNOMED CT coded data that was entered prior to that time when the concept was active (but assuming that all more recent new data entry prevents users from selecting inactive concepts)
- The active substitute declared for the inactive concept is equal to any of a limited list of very high level, abstract grouper concepts in SNOMED CT, with correspondingly limited querying value even if they were included in the full Query Table:

```
138875005 SNOMED CT Concept (SNOMED RT+CTV3) 363661006 Reason not stated concept (inactive concept)
105590001 Substance (substance) 363662004 Duplicate concept (inactive concept)
106237007 Linkage concept (linkage concept) 363663009 Outdated concept (inactive concept)
123037004 Body structure (body structure) 363664003 Erroneous concept (inactive concept)
123038009 Specimen (specimen) 370126003 Moved elsewhere (inactive concept)
243796009 Situation with explicit context (situation) 443559000 Limited status concept (inactive concept)
254291000 Staging and scales (staging scale) 91722005 Entire physical anatomical entity (body structure)
260787004 Physical object (physical object) 301857004 Finding of body region (finding)
272379006 Event (event) 118234003 Finding by site (finding)
308916002 Environment or geographical location (environment / location) 123946008 Disorder by body site (disorder)
362991000 Qualifier value (qualifier value) 362966005 Disorder of body system (disorder)
363787002 Observable entity (observable entity) 362958002 Procedure by site (procedure)
370115009 Special concept (special concept) 669771000000105 Operations by other sites or systems (procedure)
373873005 Pharmaceutical / biologic product (product) 309040001 Operations by other sites or systems (procedure)
404684003 Clinical finding (finding) 118664000 Procedure on body system (procedure)
410607006 Organism (organism) 128927009 Procedure by method (procedure)
419891008 Record artifact (record artifact) 118665007 Executable component (procedure)
48176007 Social context (social concept) 315645005 Unspecified conditions (disorder)
71388002 Procedure (procedure) 2422100250000103 SNOMED CT UK administrative concepts
301857004 Finding of body region (finding) (administrative concept)
78621006 Physical force (physical force) 386985000 Evaluation procedure (procedure)
386985000 Evaluation procedure (procedure)
```
5 Limitations

The Query Table is conceived as an optimised derivative of the information represented in SNOMED CT as historical relationships, asserting ‘semantic equivalence’ links between inactive concepts and their active counterparts (ie the blue lines in Figure 1 and Figure 2).

Provided such links exist, and are correct when they do exist, the intended effect of the Query Table is: a search that should return an active concept will also return all its inactive equivalents, and vice versa.

It is, however, known that at least some historical and taxonomic relationships represented in current SNOMED CT content are not, in fact, clinically correct. More significantly, not all valid historical relationships are actually present in current SNOMED CT content. For example, known cases of incomplete historical association data includes those many inactive concepts stated to be ambiguous but for which only one alternative active concept is specified (identifiable in the Substitutions table as those with IsAmbiguous = 1) plus some volume of ambiguous inactive codes where, although more than one active substitute is currently specified, the set of substitutes is not complete.

Of particular interest should also be those inactive concepts appearing in the UK Edition but without any active, semantically equivalent substitute declared; these can be easily identified within the Substitutions Table as those rows where OLDCONCEPTID is the same as NEWCONCEPTID. As of Aug 2020 there were 49,567 such ‘orphaned’ concepts.

Neither the Substitutions nor Query Tables can mitigate the theoretical risk of false negative reporting for ‘orphaned’ concepts.

However, the real clinical risk arising from orphaned concepts is likely to be very small: the huge majority relate to SNOMED codes corresponding to NHS Drug and Appliance Dictionary codes, which have no mapping to dm+d by deliberate design and which are therefore already out of scope for the Query Table.

This phenomenon was studied in greater depth in 2016, when a residue of 9,758 orphaned concepts not derived from the old NHD Drug and Appliance Dictionary was identified. Further analysis of this subgroup revealed that 75% had been inactivated prior to 2012 and all but 156 related to one of 18 known reasons why no active semantically equivalent concept could exist:

<table>
<thead>
<tr>
<th>REASON</th>
<th>NUM</th>
<th>REASON</th>
<th>NUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHS DRUG AND APPLIANCE DICTIONARY</td>
<td>79459</td>
<td>UNILATERAL/BILATERAL</td>
<td>34</td>
</tr>
<tr>
<td>VETERINARY MORBIDITY, ANATOMY OR PATHOGEN</td>
<td>4886</td>
<td>EXINCT CROSS-TYPE CONTENT</td>
<td>31</td>
</tr>
<tr>
<td>ORIGINAL CONCEPT AUTHORED IN ERROR</td>
<td>4034</td>
<td>LIST CLOSURE CATEGORY FROM ICD</td>
<td>31</td>
</tr>
<tr>
<td>AMBIGUOUS READ/CTV3 CODE</td>
<td>161</td>
<td>CLINICAL COURSE/ONSET</td>
<td>14</td>
</tr>
<tr>
<td>CTV3 METADATA</td>
<td>91</td>
<td>DISJUNCTION</td>
<td>14</td>
</tr>
<tr>
<td>SURGICAL APPROACHES</td>
<td>81</td>
<td>EXTERNAL ANATOMICAL FEATURES</td>
<td>14</td>
</tr>
<tr>
<td>ACTION WITH EXPLICIT DEVICE OR SITE</td>
<td>79</td>
<td>ANIMAL HAIR</td>
<td>10</td>
</tr>
<tr>
<td>VETERINARY/OBSOLETE (MICRO)ORGANISM TAXON</td>
<td>76</td>
<td>OBSOLETE NHS ORGANISATION</td>
<td>10</td>
</tr>
<tr>
<td>BRANDED FOODS</td>
<td>48</td>
<td>NAVIGATIONAL CONCEPT</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GROUPER CATEGORY FROM ICD</td>
<td>6</td>
</tr>
</tbody>
</table>

Taken together, this analysis of the timing of the inactivation and the semantic scope of the concepts involved suggests that (a) few of the codes involved could have ever plausibly been selected to capture clinical problems or interventions (or, if they were, would have constituted serious miscodes) and (b) most of the affected data is clinically no longer current.

The final residue of 156 identified in 2016 are listed below, grouped by their SNOMED CT semantic tag; a small number (listed in bold) had been identified as clearly having directly equivalent active concepts, and content change requests filed to correct those errors:

10761004|Topography not applicable (body structure)
113178003|Calamus (body structure)
13319006|Proximal interphalangeal joint structure of great toe (body structure)
181641001|Entire flexor digitorum superficialis muscle (body structure)
307209008|Grade 5b (qualifier value)
360023008|Specialized action (qualifier value)
82378008|Unlisted physical agent (qualifier value)
82975001|Denied (qualifier value)
6 File format, implementation, usage and test harness

The SNOMED CT Query Table is presented as a 3-column, TAB separated file. Rows are terminated by CR/LF combination, and the first row contains the relevant column names.

The table schema is as follows:

<table>
<thead>
<tr>
<th>COLUMN</th>
<th>LENGTH</th>
<th>TYPE / PATTERN</th>
<th>DATABASE TYPE</th>
<th>NOTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPERTYPEID</td>
<td>18</td>
<td>SCTID</td>
<td>VARCHAR (18) or BIGINT(10) NOT NULL</td>
<td>SNOMED CT Concept Identifier. Not unique</td>
</tr>
<tr>
<td>SUBTYPEID</td>
<td>18</td>
<td>SCTID</td>
<td>VARCHAR (18) or BIGINT(10) NOT NULL</td>
<td>SNOMED CT Concept Identifier. Not unique</td>
</tr>
<tr>
<td>PROVENANCE</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

6.1 Column Details

SUPERTYPEID

The 64-bit integer identifier for a SNOMED CT concept (up to 18 digits). SNOMED Ids may be stored as VARCHAR(18), but indexing and lookup operations typically run faster when stored as BIGINT(10).

SUBTYPEID

The 64-bit integer identifier for a SNOMED CT concept (up to 18 digits) that, for the purposes of taxonomic querying, is to be treated as though it were a taxonomic descendent of the concept referenced as SUPERTYPEID.

SNOMED Ids may be stored as VARCHAR(18), but indexing and lookup operations typically run faster when stored as BIGINT(10).

PROVENANCE

A single digit value, encoding the provenance of the inference that SUBTYPEID is a descendent of SUPERTYPEID; currently in range 0-3. Other values may be introduced as a result of planned content improvements:

<table>
<thead>
<tr>
<th>PROVENANCE VALUE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SUBTYPEID is related to SUPERTYPEID exclusively by means of regular taxonomic inheritance plus one or more of:</td>
</tr>
<tr>
<td></td>
<td>1. SAME_AS, REPLACED_BY or MOVED_FROM historical relationships</td>
</tr>
<tr>
<td></td>
<td>2. Default context equivalence</td>
</tr>
<tr>
<td></td>
<td>3. CTV3/READ2 Navigational Subset inference</td>
</tr>
<tr>
<td></td>
<td>(including possibly by chains of indefinitely many relationships exclusively of these types)</td>
</tr>
<tr>
<td></td>
<td>These types of relationships are interpreted as unambiguously implying true</td>
</tr>
</tbody>
</table>
concept identity between pairs of concepts ie *everything* true of one is also true of the other.

1 SUBTYPEID is related to SUPERTYPEID via regular taxonomic inheritance *plus* at least one MAY_BE relationship, but where every inactive subject of the MAY_BE relationships involved has only one candidate active SNOMED CT conceptId substitute.

Theoretically, ambiguous inactive concepts for which only one active substitute is specified *should* imply that SNOMED CT no longer offers a means to encode any other possible interpretations of the original code and, therefore, no false positive substitutes exist to be returned.

Practically, however, this phenomenon may also occur when SNOMED CT’s set of historical MAY_BE relationships is incomplete.

Accordingly, whilst it is more likely to be the case that the designated substitute is a true positive, and no false positives or false negatives exist, users’ attention may reasonably be drawn to search results obtained using these rows.

2 SUBTYPEID is related to SUPERTYPEID by means of regular taxonomic inheritance *plus* at least one WAS_A relationship, so that everything that is true of SUPERTYPEID is also true of SUBTYPEID but the reverse is not known to be true. For example, the ancestors of SUPERTYPEID are also ancestors of SUBTYPEID (but SUBTYPEID may also have other ancestors). However, the descendants of SUPERTYPEID may not be also descendants of SUBTYPEID.

3 SUBTYPEID is related to SUPERTYPEID by means of regular taxonomic inheritance *plus* at least one overtly ambiguous MAY_BE historical relationship that is associated with an inactive concept for which more than one possible active substitute is given.

Results obtained using these rows will therefore almost certainly include false positives, and so users should normally be alerted.
6.2 Implementation and usage

Recap: Transitive Closure tables

In order to understand how the Query Table should be implemented, it is first necessary to revisit the design and purpose of an ordinary Transitive Closure table.

An ordinary Transitive Closure Table lists all possible pairwise supertype:subtype relationships that can be inferred from the set of 116680003 Is a relationships listed in the union of all sct2_relationship_snapshot files loaded as part of some data configuration.

The Transitive Closure table is therefore typically used as the means by which an Expression Constraint Language fragment of the general form $<<X$ can be expanded into a full enumeration of all descendants of $X$. The required expansion query is simple:

\[
\text{SELECT SUBTYPE FROM TRANSITIVECLOSURE WHERE SUPERTYPE = X;}
\]

However, this method of expanding $<<X$ only produces a clinically plausible result when $X$ is itself active. All inactive $X$ have zero descendants, and therefore asking only a conventional transitive closure table for all SNOMED codes that correspond to “kinds of” of any inactive concept will universally say that there are none even though clinically this is obviously untrue. For example, since January 2019, all SNOMED CT Transitive Closure tables will report that SNOMED contains zero other codes that are “kinds of” 262536007|Superficial abrasion|, despite there being many clinically relevant codes plain to see under 399963005|Abrasion|.

In fact, however, results returned by the Transitive Closure table for the “kinds of” even active codes are also always incomplete with respect to never containing any inactive but clinically relevant codes such as - using again the $<<262536007|Superficial abrasion|$ example above - 157467002|Abrasion of upper limb|. This occurs because all inactive $X$ also have zero ancestors and, therefore, no active $X$ can have an inactive descendent.

Implementation

The Query Table is therefore conceived primarily as an extension – extra rows – for a normal Transitive Closure table, and where the addition of these extra supertype:subtype rows will substantially correct the issues outlined above. It is computed by traversing both the 116680003 Is a relationships and the set of historical associations in SNOMED - both the black and blue lines in Figure 2 - but with the latter extracted from a History Substitutions table in which the complexity of the historical associations published in SNOMED out-of-the-box has already been analysed and summarised.

The final published Query Table content has two important characteristics:

1. It contains only those additional supertype:subtype inferences not also found in an ordinary Transitive Closure table. For this reason, expanding $<<X$ against the Query Table only will also always yield a clinically (and typically more obviously) incomplete result.

2. It assumes that the Transitive Closure table it is paired with will have been constructed to return all codes as their own subtype, whether active or inactive. This reflexive relationship for inactive codes is not published in the Query Table.

Critically, because the Query Table is an extension to rather than a substitute for a Transitive Closure table, $<<X$ should always be expanded only against the union of the Transitive and Query Tables. The result obtained will include both active and inactive clinically expected results, regardless of whether or not $X$ itself was active or inactive. It is therefore never necessary to first test whether or $X$ is active before deciding how to expand $<<X$. 
An implementation choice concerns whether to maintain the tables separately, querying them independently and then merging their responses in software. A viable alternative could be to concatenate them at load time into a single table: this can be achieved by adding a PROVENANCE column to the normally 2-column Transitive Closure table and populating it with e.g. either 0 or -1, before finally appending all rows in the published Query Table product.

The obvious benefit of simple concatenation as one table is somewhat simpler querying code; disbenefits include that the resulting table will be very large (currently approaching 28 million rows), and it may be more difficult to offer the ability to switch elements of Query Table use off and on, if the use case requires that flexibility.

**Tuning: use of PROVENANCE field**

The purpose of the PROVENANCE field is to allow queries directed at the Query Table element to be “tuned” according to the level of false negative and false positive reporting that the particular use case will tolerate, or the balance between them it wishes to strike.

Most taxonomic query implementations, regardless of use case, will normally and legitimately *always* include lookup on rows where ISAMBIGUOUS=0, because this indicates where the supertype:subtype inference is fundamentally based only upon unambiguous and semantically very strong statements of simple, global identifier equivalence. Users therefore need not usually be notified of any results obtained that are dependent on such rows because there is typically zero risk of false positive reporting. However, if only PROVENANCE=0 rows are polled there will remain a residual risk of false negative reporting in respect of inactive codes with weaker or frankly ambiguous statements of equivalence to active codes. Exceptions to this general rule are primarily specialist use cases where a technical requirement is that expansions of $<<X$ should *not* contain inactive codes.

Lookup on rows where ISAMBIGUOUS = 2 (ie supertype:subtype inferences that depend on traversing a WAS-A relationship) do not normally entail any risk of false positives, but users may well benefit from being notified that the risk of false negatives is unlikely to have been fully mitigated.

Queries executed in an environment where false positives can be tolerated (e.g. case finding where a second validation by human inspection step will occur) can also normally include lookup of rows where ISAMBIGUOUS=1 or 3, but that portion of the result set obtained by this route should be highlighted to the user tasked with weeding out the false positives.

In safety critical situations where false positives can not be tolerated – e.g. when running reports to automatically enrol patients for interventions – but where no manual review to weed them out can be implemented, the ISAMBIGUOUS=1 or 3 rows should usually NOT be consulted. In some settings it may however be judged acceptable to include the ISAMBIGUOUS=1 rows without human review.

### 6.3 Sample implementation/test harness and SQL

Stylised SQL for expanding ECL expressions of the general form $<<X$ could therefore be:

```sql
SET @Probe1 = X;
SET @TUNING = 1;
SELECT DISTINCT Include.SUBTYPEID
FROM (  
    SELECT qi.SUBTYPEID, qi.PROVENANCE FROM SCT_QUERY AS qi WHERE qi.SUPERTYPEID = @Probe1
    UNION (SELECT tci.SUBTYPEID, 0 AS PROVENANCE FROM SCT_TC AS tci WHERE tci.SUPERTYPEID = @Probe1)
    UNION (SELECT ci.id,ci.id,0 FROM SCT_CONCEPT ci WHERE ci.id = @Probe1)  
) AS INCLUDE
WHERE Include.PROVENANCE <= @TUNING;
```
…and therefore \((<<X\ MINUS\ <<Y)\) becomes:

```
SET @Probe1 = X;
SET @Probe2 = Y;
SET @TUNING = 1;

SELECT DISTINCT Include.SUBTYPEID FROM
  (SELECT qi.SUBTYPEID, qi.PROVENANCE FROM SCT_QUERY AS qi WHERE qi.SUPERTYPEID = @Probe1
  UNION (SELECT tci.SUBTYPEID, 0 AS PROVENANCE FROM SCT_TC AS tci WHERE tci.SUPERTYPEID = @Probe1)
  UNION (SELECT ci.id, ci.id, 0 FROM SCT_CONCEPT ci WHERE ci.id = @Probe1)) AS INCLUDE
LEFT JOIN
  (SELECT qx.SUBTYPEID, qx.PROVENANCE FROM SCT_QUERY AS qx WHERE qx.SUPERTYPEID = @Probe2
  UNION (SELECT tcx.SUBTYPEID, 0 AS PROVENANCE FROM SCT_TC AS tcx WHERE tcx.SUPERTYPEID = @Probe2)
  UNION (SELECT id, id, 0 FROM SCT_CONCEPT_FSN cx WHERE cx.id = @Probe2)) AS EXCLUDE
ON (Include.SUBTYPEID = Exclude.SUBTYPEID AND Exclude.PROVENANCE <= @TUNING)
WHERE Exclude.SUBTYPEID IS NULL AND Include.PROVENANCE <= @TUNING;
```

A small Query Table test harness implementation is included within the Query Table distribution. It takes the form of a simple table of simulated coded EPR data, and a MySQL script file to load that data and then query it both with and without the Query Table.

The SQL script file requires the prior existence of a ‘snomed’ schema containing at least the following three tables, as named below:

1. rf2_concept_sp - a recent snapshot of the rf2_concept table from the RF2 distribution of the UK Edition of SNOMED CT

2. TRANSITIVECLOSURE - a precomputed transitive closure table for the same edition of SNOMED CT (but where this returns X as SubtypeOf X for all X, whether active or inactive)

3. QUERYTABLE - the Query Table as distributed

Further outline instructions on how to set up and then use the simple test harness are provided at the top of the same SQL script file.